

Laws and Theories in Quantitative Linguistics

Peter Meyer¹

Abstract. According to a widespread conception, quantitative linguistics will eventually be able to explain empirical quantitative findings (such as Zipf's Law) by deriving them from highly general stochastic linguistic 'laws' that are assumed to be part of a general theory of human language (cf. Best (1999) for a summary of possible theoretical positions). Due to their formal proximity to methods used in the so-called exact sciences, theoretical explanations of this kind are assumed to be superior to the supposedly descriptive-only approaches of linguistic structuralism and its successors. In this paper I shall try to argue that on close inspection such claims turn out to be highly problematic, both on linguistic and on science-theoretical grounds.

Keywords: Zipf, language laws, *ceteris paribus* laws, emergence, complexity theory, explanation, science theory, word length, Menzerath

0. Introduction

Quantitative linguistics (henceforth, QL) as understood in the following considerations is concerned with accounting for quantifiable and measurable linguistic phenomena in terms of mathematical models such as curves, probability distributions, time series and the like. The mathematical formulas employed are attributed the status of scientific laws to the extent that they are deducible from very general principles or 'axioms' and are thus firmly integrated into some nomological network. Inasmuch as investigations in QL fulfil these basic requirements, they are believed to be paradigm cases, indeed the very first of their kind in the whole history of linguistics, of empirical scientific theories in a narrow, science-theoretically justified sense, as opposed to purely descriptive and taxonomical approaches in the traditional, 'qualitative' branches of linguistics, the latter being therefore charged with failing to attain the high methodological standards of the natural sciences (Altmann 2002, Altmann, Lehfeldt 2002 provide sufficient documentation for these claims). Likewise, G. K. Zipf, one of the founders of the contemporary methodology of QL, has lately been advanced to being "the first language theoretician", even the Newton of a new kind of theoretical, namely, 'Zipfian' linguistics (Altmann 2002, 22; 25).

The pungent methodological criticism of 'traditional', 'qualitative' linguistic approaches advanced in the QL literature constitutes a good starting point for the considerations to follow, since the basic science-theoretical and linguistic assumptions on which most work in QL is founded are to be understood as a corollary of this rejection of qualitative and descriptive methods. The criticism in question is, as I would like to show, based on a fairly restrictive and questionable notion of scientific theory. Thus for Altmann (2002, 19), 'qualitative' accounts for linguistic phenomena necessarily remain at a "proto-scientific level" because they cannot "satisfy the claims of natural sciences". In remarks such as these a certain normative science-theoretical stance becomes apparent, a stance according to which theories

¹ Address correspondence to: Peter Meyer, Seminar für Slavische Philologie, Georg-August-Universität Göttingen, Humboldtallee 19, D-37073 Göttingen; e-mail: meyer-peter@gmx.de

that deserve to be called empirical and scientific have to be modeled following the example set by the natural sciences, or strictly speaking, by axiomatized fundamental physics only. This attitude seems to date back to at least the days of the Logical Positivists and may firmly be said to be superseded by a much more differentiated contemporary discussion, represented, e.g., in textbooks such as (Balzer 1997). From the ample literature on the subject, it is usually only Mario Bunge whose works are quoted in support. Bunge (1995, 3) demands that linguistics confine its methods to quantifiable magnitudes that are referred to by the laws of a scientific theory proper:

Can every feature be quantitated, that is, turned into a magnitude? I submit that only one property is, with all certainty, intrinsically qualitative, namely existence. I also submit that in every other case quantitation depends exclusively on our ability and interest, so that in the face of failure to quantitate we should suspend judgement and encourage others to try.

All this amounts to is an ill-founded profession of faith² that is refuted not only by well-established scientific practice in the social sciences and in the humanities. Recent science-theoretical surveys abound with reconstructions of purely qualitative theories (see, e.g., Balzer 1997 or Balzer, Moulines, Sneed 1987). Linguistics is a field that provides some paradigm cases of successful scientific modeling that fully meet explanatory requirements. The reconstructive methodology of historical linguistics is a case in point, providing, at least to a certain extent, even the possibility of prediction. There is no good reason not to call, say, some historical grammar of the Indo-European languages such as the one summarized in Beekes (1995) a partial (and admittedly not fully explicit)³ *theory* of the historical development and relationship between the languages in question. The hypothetico-deductive method, which is often seen as a cornerstone of modern science, has undoubtedly become a paramount methodological instrument of contemporary ‘qualitative’ linguistics, mostly due to the influence of Chomsky’s writings whose theory of human language, at least in its modern principles-and-parameters guise, belongs among the most convincing examples of ‘non-numerical’ theories, notwithstanding logically independent quarrels about whether it is good or ‘approximately true’ theory.⁴ Even any good traditional grammar or dictionary of a language is a theory of that language, however incomplete, implicit and embryonic. Many concepts of such language theories are, *pace* Bunge, *inherently* qualitative, presupposing a yes-no decision that cannot reasonably be made ‘fuzzy’ or otherwise turned into a quantitative magnitude. A certain lexeme or phoneme either appears in this or that utterance token or it does not appear. *Tertium non datur*: A lexeme cannot be said to appear in the utterance “to a degree of 70%”, although it would of course be possible to say, e.g., that “70% of all speakers think that lexeme A occurs in this utterance”. But even in such a quantitative judgment a class of discretely individuated lexemes (among them, lexeme A) is already inevitably presupposed. As we shall see in section 6, it is precisely (and, from the point of view of QL,

² Ironically, the only qualitative property Bunge accepts at all, existence, is not regarded as a property at all by the vast majority of philosophers since Kant.

³ It is important to notice that there are fundamental limits to explicitness even in highly formalizable theories of the ‘exact sciences’. To give but one example, according to one of the most eminent researchers in this field of the philosophy of science, a deductive axiomatization of experimental physics is impossible (Suppes 1998).

⁴ The theory of language advanced by Chomsky and his followers tends to be misdescribed in QL writings as a mere formalism to express language-specific rules of grammar that have been found by inductive generalization. As a matter of fact, however, contemporary generative grammar takes as its axioms, amongst other things, the uniform initial state of an assumed human language faculty and *deductively derives* from these axioms, amongst other things, predictions as to the implicit knowledge of an adult native speaker under given experiential boundary conditions. This implicit knowledge can be *tested* using empirical methods. Hence, modern generative methodology fulfils all requirements for theoryhood typically formulated in QL. See Chomsky (2000) for an accessible recent presentation.

ironically) modern complexity theory and related developments in abstract evolution theory that provide some deep arguments in favor of the inevitability and irreducibility of qualitative, descriptive and functional accounts of certain complex systems.

Bunge's demand that all linguistic properties be quantifiable is an outgrowth of a certain general view on what characterizes a genuine scientific theory. A concise recent summary of this view can be found in (Altmann, Lehfeldt [to appear]). According to the authors, the notions of 'explanation' and 'law' are adequate only in case the fundamental statements of the theory in question are of a *dynamic* nature in an abstract, quantitative sense, that is, representable as difference or differential equations or as self-regulation schemata in some sort of systems theory. Once again, we are left with some sort of metatheoretical *credo* that leaves open why *this* should be the only *modus operandi* allowed in modern science, let alone linguistics, particularly as the immense complexity of social and neurophysiological processes that jointly underlie the dynamics of language make it seem rather implausible that this dynamics can be modeled in any *interesting* way in terms of, say, a bunch of differential or equilibrium equations. The *ultima ratio* behind the methodological propensities typical of QL proponents seems to be the irresistible attraction exerted by the now-fashionable scientific paradigms of chaos and complexity theory; see below for some critical evaluation. For the time being, the only thing that can justifiably be said with respect to the methodology of QL is that it is a way of looking at language that is *complementary* to traditional approaches and cannot, for this reason, be translated into the conceptual apparatus of the latter or *vice versa*.

In a number of publications (e.g., Altmann, Lehfeldt 2002) the notion of 'theory' is defined in contradistinction to mere inductive generalization as allegedly offered by traditional, 'qualitative' approaches. To begin with, it must be stressed again that this allegation ignores the role of deductive-nomological explanations and of the hypothetico-deductive method in contemporary linguistics. Many of Chomsky's writings sound remarkably similar to recent contributions to QL in rejecting arbitrary inventorization of data and superficial empirical generalizations in favor of deep and unified explanatory principles from which empirical generalizations can be deduced; for a succinct early statement cf. Chomsky (1978). In this paper, I will assess several criteria proposed in the QL literature that are used to justify the status of a scientific theory for QL models. These criteria may be summarized as follows. The backbone of any scientific theory proper is formed by *laws*. Typical empirical hypotheses of QL (such as, say, distribution of word length in texts) are indeed laws in a strict science-theoretical sense since they are embedded in a nomological network, i.e. they are *deducible* from underlying postulates or 'axioms'; and, by virtue of referring to measurable quantities, these laws are subject to empirical *confirmation* or *disconfirmation*.

It is the main objective of this paper to maintain that the quantitative regularities discovered so far in QL do *not* pass as law-like statements, that is, as analogues to what qualifies as 'laws' in the natural sciences, particularly in fundamental physics. As a consequence, *explanations* for these regularities (in a science-theoretically established sense of 'explanation') are still wanting. It is, however, a delicate task to assess the possible impact and importance of the claims just put forth on QL research work. What is not claimed here is that the results obtained so far in QL are empirically or theoretically void. Nor will it be the purpose of the following remarks to impose certain normative science-theoretical requirements on QL. Quite to the contrary I would like to suggest that QL faces the danger of being caught up in a false picture of what constitutes 'real' science, a picture that could serve as a problematic guideline to further research work in drawing the wrong dividing line between what should be considered 'good' and 'bad' questions in QL.

Most of the critical reflections put forward in the present article are not novel. As early as in 1959, a stimulating review article on the book *Logique, langage et théorie de l'information* by B. Mandelbrot, L. Apostel, and A. Morf (Lees 1959) succinctly put its fingers on many of

the conceptual problems and inherent limitations of a quantitative treatment of language. I shall permit myself to quote some of Lees's still relevant remarks in footnotes. In addition, it must be emphasized that some of the chief architects of contemporary QL are very explicitly aware of unresolved theoretical problems in the discipline (see esp. Grotjahn, Altmann 1993; Altmann 1999; Altmann 2002).

1. A case study: word length

For the purposes of our discussion I will take the theoretical treatment of word length distribution in German texts presented in Altmann, Best (1996) as a typical example of how a certain quantitative law-like statement can be taken to explain certain statistical regularities. Similar examples would of course have to be discussed for other putative laws of QL. I will simply assume here that the foundational problems observed with respect to my example case also arise in the context of other would-be QL laws, for analogous reasons.

The fact that the negative binomial distribution can be fitted well to the empirical distribution of word length (measured as number of syllables) in a wide variety of German texts is explained in this paper by stipulating an underlying self-regulative mechanism that consists in mutual influence between neighboring word length classes. In accordance with the general framework proposed in Wimmer et al. (1994), it is assumed that this influence implies a proportionality relation between neighboring classes:

$$(1) \quad P_x = g(x)P_{x-1}$$

(1) is assumed to be the underlying *law-like principle*⁵ that governs word length distribution in general. $g(x)$ is a language-specific proportionality function that, for the German texts in question, is assumed to be representable as

$$(2) \quad g(x) = \frac{a + bx}{cx}.$$

Here, a is taken to represent something like the length-invariant part of the German lexicon, whereas b is an author-specific modification factor (the author chooses to employ shorter or longer words, according to stylistic and other needs) and c 'stands for' the communicative interests of potential text recipients (such as minimizing the effort of decoding a given message). From (1) and (2), an explicit representation of P_x can be derived. The deduction yields a family of univariate discrete probability distributions with two parameters, namely, that of the negative binomial distribution. Now this family of distributions can indeed be fitted to the German text data in question. Thus, the underlying principles seem to receive empirical confirmation. In what follows, we will examine in turn the different parts of the methodology just sketched.

⁵ Thus, Wimmer, Altmann (1994) write: "We consider statement (1) as a law-like hypothesis since it fulfills the requirements put on laws ..., above all generality, systemicity and confirmation. Nevertheless it is merely a skeleton that must be filled with flesh taken from languages, genres and authors, all of them bringing different boundary and subsidiary conditions which can vary in the life of language or of an author. Thus no *specific* formula following from (1) holds eternally for all languages or even one language."

2. The notion of law

It will be useful to compile some definitional statements about ‘lawlikeness’ as found in the QL literature. Altmann (1993) writes, quoting Bunge in support:

“Only syntactically well-formed, semantically meaningful general statements that are empirically testable, not including observational concepts, stating something about invariances and going beyond our present knowledge should be considered as hypotheses. If a hypothesis is derived from assumptions (axioms) or from a theory, if it is corroborated by an empirical test and if it can be connected with other similar statements (systematized), then we can call it a law.”

In what follows I will examine in particular whether QL hypotheses are indeed derivable from assumptions or axioms (section 3) and to what extent they can be corroborated empirically (section 4).

Empirical confirmation of QL hypotheses such as (1) above is impeded by the possibility of ‘exceptions’ that have to be taken account of in some way. Altmann, Erat, Hřebíček 1996 write on behalf of the empirical validity of (1):

We can consider the probability distribution as a kind of attractor, a form existing in every language – i.e. existing unconsciously in the text users – to which the empirical distributions of the given variable tend. Of course, there can be a number of different attractors exerting their impact on individual writers or on individual genres, and each of them can evolve in the course of time. As a matter of fact, formula (1) merely represents a mechanism which can take into account a number of boundary or subsidiary conditions, as is the case in natural laws, too. In practice, if a text deviates from the supposed attractor, we say that it wanders to another attractor, which is quite a normal circumstance in the life of a text producer. This wandering can be expressed in different ways (cf. Wimmer et al. 1994), e.g. in the modification of $g(x)$, in the increase of the order of the difference equation (1), in the modification of the individual frequency classes, in the mixing, compounding or convolution of probability distributions, etc.

Section 4 will also criticize the strategy hinted at in this quotation, namely of attributing QL hypotheses the status of *ceteris paribus* laws that are effective only under certain boundary conditions that cannot be listed explicitly and exhaustively.

In addition, the above quote also refers to underlying ‘mechanisms’ that are supposed to be described by the laws of a theory.⁶ Section 5 will deal with some of the stipulated language mechanisms that are assumed to ‘generate’ the quantitative regularities that have been observed so far.

3. Deriving laws from axioms

Turning back to our case study, it must be stressed at the outset that a purely mathematical deduction of the negative binomial distribution from the difference equation (1) plus specification (2) does not provide us with a theoretical explanation of the distribution, since it does not embed it in a nomological network that has an *independent justification*. (1) and (2) are nothing but a mathematically equivalent reformulation of the probability distribution.⁷

⁶ Thus, Altmann (1993) writes: “**Laws** are statements about mechanisms which generate observable phenomena.”

⁷ Altmann (1980) acknowledges this point, when he comments on the derivation of Menzerath’s law from a simple differential equation: “The derivation from a differential equation is not sufficient in order to award the statement (4) [Menzerath’s law in a quantitative formulation, P.M.] the status of a law. It remains a theoretically not fully validated hypothesis as long as it is not set in relation to other laws of language, i.e. until it is

What we need is a testable criterion that tells us when it is appropriate to assume that (1) and (2) hold.

Recent efforts in QL (cf. Wimmer, Altmann 2003) concentrate on finding a ‘unified derivation’ of linguistic laws. This amounts to finding a very general class of difference / differential equations⁸ from which all those formulas that have been employed in descriptive QL models so far can be derived. While the principle idea of the authors consists in epistemically integrating disparate research domains under the heading of a new ‘supertheory’ that contains the old particular ones as special cases, we are eventually left with the observation that different mathematical *formulas* employed in QL – representing variously probability masses or densities or function curves – may be transformed into some sort of very general ‘canonical form’. There are no good reasons to assume that this purely formal analogy between extremely different formulas used in wildly disparate *interpretations* (as probability densities, as functions etc.) has a deeper reason connected somehow with (universal) properties of *human language*. All we get is a purely mathematical observation that has not yet any clear implications for the phenomena described with the aid of the respective formulas.

To sum up: The observation that the family of probability distributions defined by (1) and (2) can be used to ‘model’ word length in a variety of texts does not make (1) (or, for that matter, (1) *cum* (2)) a law statement, let alone a deductive-nomological explanation of observed word length distributions. For this to be the case, we would need some further justification for positing something like (1) as the general principle governing word length in human language. Interestingly, Wimmer et al. (1994) hint at some ideas to give (1) some initial plausibility:

We assume that the various word length classes do not evolve independently of each other. If there is a gradual increase of disyllabic words in a language evolving from monosyllabism to polysyllabism ..., this occurs as a function of the number of new meanings that must be coded and of the degree of polysemy and redundancy in the class of monosyllabic words. If the redundancy in this class has reached a critical level, the equilibrium must be restored by means of functional equivalents, e.g. tones, new phonemes, extension of phoneme distribution, variation of word length, etc. If a language has recourse to polysyllabism, the class of disyllabic words must necessarily be made proportional to that of monosyllables, i.e. in probabilistic terms

$$P_2 \sim P_1.$$

If a language is no longer restricted to monosyllabism and polysyllabic words are introduced, then self-regulation comes into play and controls the whole frequency structure of word length.

In the first step, i.e. when disyllabic words are introduced, proportionality can be considered as constant, i.e. $P_2 = aP_1$. If longer words come into existence, constant proportionality will be replaced by a function of length $g(x)$. We thus obtain the basic formula

$$(1) \quad P_x = g(x)P_{x-1}.$$

Even for the simple case of a ‘monosyllabic’ language developing disyllabic words, however, the authors’ argument is far from clear. For a given text in the language in question, two probabilities must be posited, namely, the probabilities P_1 and P_2 that a given word form token in the text consists of one and two syllables respectively, where $P_1 + P_2 = 1$, hence the

incorporated into a system of laws. Such a system does not exist at present, we merely suspect that it is somehow connected with the principle of least effort or with some not yet known principle of balance recompensating lengthening on one hand with shortening on the other.”

⁸ The approach runs roughly as follows: The relative rate of change of a variable Y is taken to be dependent on the rate of change of one other independent variable X. The latter is itself controlled by different powers of X that are associated with different multiplicative factors.

proportionality factor a is already given as $a = \frac{1}{P_1} - 1$. Now it is obvious that one cannot say

of two *numerical values* that they stand in a proportionality relation to each others since only *functions* can be proportional to each other ($f(x)$ and $g(x)$ are proportional to each other iff $g(x) = \text{const.} \cdot f(x)$). Thus, the proportionality posited by Wimmer et al. makes sense only inasmuch as word length probability is considered a function of some independent variable, e.g., text length. Empirical results show that the proportionality factor will vary from text to text, a point also stressed by the authors. Hence, the proportionality factor as assumed by the authors implies some kind of counterfactual conditional of the following kind: “Had we examined another text that is in all relevant respects similar to *this* text, we would have obtained the very same ratio of disyllabic to monosyllabic words.” As long as we do not have any non-circular account of what the phrase “in all relevant respects” means here, the conditional is virtually devoid of meaning. And even if we had such an account we would still be in need of a scientifically valid justification for the quantitative principle (1) besides the air of vague qualitative plausibility. Hence, the ‘proportionality principle’ (1) remains unjustified in a strict sense of the word even for the simple case of a language restricted to two word length classes and $g(x) = \text{const.}$, since we cannot figure out what proportionality should mean here at all. Analogously, for the general case, in which we do not have a proportionality *constant* but a ‘proportionality function’, principle (1) becomes literally tautological because, trivially, for any discrete univariate probability distribution $P(x)$, (1) can be made true by defining a ‘proportionality function’ $g(x)$ that fulfills the equation $g(x) = \frac{P(x)}{P(x-1)}$. It is clear

that principle (1) gains empirical character only by virtue of specifying $g(x)$. However, we are not offered any theoretically well-founded restrictions on to what class of functions $g(x)$ should belong, only some inductive evidence on what functions ‘worked well’ in past investigations, that is, have led to a good fit for a reasonable amount of texts. Nor do we possess any criterion for predicting which selection among a set of ‘approved’ functions will do well for a newly investigated text. In other words, (1) cannot be falsified; hence, it is not an empirical principle in any sense and, therefore, is not capable of being an ‘axiom’ from which theorems of word length distribution could be derived.

The reader might wonder whether the preceding remarks do not miss the very point of theoretical reasoning in QL since it is wrong to require that the ‘principles’ from which we deduce testable theorems be themselves deducible or justifiable in terms of yet other principles or laws. Naturally, or so QL adherents may argue, at some point deduction must come to an end and we arrive at the ‘first principles’, that is, the ‘axioms’, of the theory, for which further justification is neither possible nor required. The extremely successful foundational equations of, say, Newtonian or quantum mechanics are far from intuitive plausibility (for the less obvious case of Newton’s second law see Weizsäcker 1985), but of course nobody would reject them for this reason. But why should anyone wish to embrace Newton’s second law as an axiom of classical mechanics but nevertheless deny some explicit version of (1) or ‘Zipf’s law’ or ‘Menzerath’s law’ the status of an axiom of QL, although in both cases empirical confirmation of statements (theorems) deduced from the alleged laws is indeed possible? The difference between the two cases is connected with the important fact, overlooked by classical Logical Empiricism, that not every formal deductive system is a theory, let alone an empirical theory. In the case of Newtonian mechanics we “know” which empirical (real-life) systems Newton’s three laws should be applicable to, that is, we are able to specify, though in a necessarily pragmatic yet non-circular fashion, the so-called ‘intended systems’ of Newtonian mechanics. The specification of the set of intended systems is an essential constituent part of any empirical scientific theory, besides the formally defined

‘models’ or ‘axioms’ themselves.⁹ The point is trivial enough to be restatable in any science-theoretical framework: We must have some criterion that tells us which empirical phenomena the theory ‘talks about’, makes predictions of etc. This criterion must be independent of the specification of the theory’s axioms.¹⁰ Otherwise, astrology would count among the respectable scientific theories for the sole reason that its applicability could be restricted *post hoc* to those cases where prediction turned out to be successful. As far as our principle (1) (where $g(x)$ must be assumed to belong to a family of previously specified functions) is concerned, there is no clearly defined class of intended systems for it. All we can point to is statements about some families of probability distributions that ‘work well’ for word length in a vaguely specified range of texts. We have no idea why (1) – in a version specified for $g(x)$ – cannot be applied to this or that text; prediction is impossible. So all we can safely say is that (1) holds whenever it is found to hold. In this respect at least, QL does not yet fare much better than astrology.

The problem of parameter interpretation looms large in the above example just as elsewhere (cf. Altmann 2002). The parameters that appear in the various ‘proportionality functions’ proposed so far in the literature suffer from a complete lack of interpretability; they are just numbers obtained by fitting the model function class to the data at hand and vary from text to text without being predictable or connectable to other empirical statements about the texts in question. To be sure, interpretations have been proposed but they are plainly *ad hoc* and not susceptible to any sort of confirmation. In the interpretation proposed by Wimmer et al. (1994), the parameters involved come to be loosely associated with Zipfian ‘forces’. None of the three postulated ‘factors’ is measurable (and no method of measuring seems to be forthcoming either), so empirical confirmation of the interpretations is impossible. Note that the authors do not even have a proper justification for associating the three factors assumed with the three parameters in the way they actually do. Why should parameter c ‘stand for’ the communicative interests of communicants? Increasing c indeed shifts the probability distribution toward smaller average word length. This seems to be the motivation behind the interpretation associated with c , since short word lengths will be favoured on the production side of the communication process. But, of course, we might as well assume a or b to be the parameter associated with the factor in question, as long as we say that a (or b) is inversely proportional to the communicative interests of speech producers. We must conclude that the interpretations suggested so far are but a ‘fifth wheel’, an ornament that plays no empirically testable role within the theoretical apparatus at all.

- Note that parameter interpretability does not imply some obscure requirement to the effect that the parameters correspond to directly observable magnitudes. A (variable) parameter may be said to be interpretable just in case there is *another* theory or law that makes an *independent* statement about the values the parameter may have. In other words, the parameter must appear in at least *two logically independent law statements*: The ‘nomological network’ must be tight enough to avoid the danger of immunization against falsifiability.

⁹ See Balzer, Moulines, Sneed (1987) for details concerning the concept of ‘intended system’ introduced here. Note that the point I want to make here does not hinge upon selecting the particular ‘non-statement’ or ‘structuralist’ science-theoretical framework as presented in these two books. The authors use the term ‘intended applications’ for what I, following Balzer (1997), prefer to call ‘intended systems’ here.

¹⁰ The range of intended systems should not be thought of as a neatly pre-defined set. As science progresses, new candidates for intended systems may be discovered. In this case, the axioms of the theory indeed define necessary conditions for candidatehood, a phenomenon called ‘(partial) autodetermination’ in the structuralist approach. This must not be taken to imply that, at least in some cases, no independent specification of a class of intended systems is necessary since autodetermination is a phenomenon of theory development through time. Synchronically, criteria for recognizing the currently agreed-upon intended systems of a theory must be sufficiently clear-cut to meet with consensus of the scientific community.

However, it is precisely this nomological network that is not yet in sight for QL. Lack of interpretable parameters turns out to be yet another indication for lack of theory.

4. Inductive corroboration (testability)

The stochastic hypotheses advanced in QL cannot be confirmed directly but only disconfirmed. Usually standard Neyman-Pearson hypothesis testing is adduced to show whether linguistic data, say, the distribution of word length in a corpus of texts, is compatible with a certain hypothesis. There are difficult problems with this type of statistical confirmation, many of which are discussed at length in Grotjahn, Altmann (1993). The main problem is that if fitting of the proposed probability distribution or curve is successful, we have only shown that there is no good empirical reason to *reject* the hypothesis: *If* the hypothesis were correct, *then* the observed discrepancy between the actual data and the expected values according to the hypothesis would not be too improbable. Usually, a large number of quantitative models can be applied successfully in the sense that fitting is possible. The difficult question is to select the ‘right’ model; it can only be solved using *theoretical* considerations that are still absent because we have no principled methodology for comparing different models or different ‘derivations’ of the same model.

In order to test “general hypotheses in which no observational concepts (such as sound, syllable, word) occur” (Altmann 1980), it is necessary to use ‘observational concepts’, that is, theoretical terms as used in qualitative descriptive linguistics. In investigations on word length, we need *operational criteria* that individuate words and tell us how long they are (as measured in syllables or phonemes). It is standard practice in QL to use whatever rough-and-ready criteria are at hand (take words to be sequences of letters between blanks etc.), as long as the procedure chosen leads to statistically significant results. In its contemporary guise, QL has no choice here, since there are no known ‘derivations’ of observed quantitative regularities that would somehow theoretically reflect the *nature* of the qualitative concepts presupposed, i.e., the *role* they have to play in the qualitative theories that give them their meanings. It is very difficult to see whether such derivations would be possible at all since QL is not an ‘autonomous’ linguistic subdiscipline in the sense of Itkonen (1983); it presupposes *with conceptual necessity* an ‘autonomous’, i.e. traditional and qualitative, description of linguistic utterances. It is important to understand why the stochastic methods of QL cannot *replace* a qualitative treatment of notions such as ‘word’ or ‘syllable’. Standard linguistic notions are *token-based*, that is, they presuppose the possibility of deciding for any given *token utterance* of which words, syllables, etc. it consists. This decision hinges upon the theoretical role of the notions employed, that is, their role within the qualitative theory of that language. QL notions can only be used to make assertions about statistical populations of utterances and do thus not have the capacity of making theoretically relevant descriptive decisions on single token utterances. Hence, even if a certain qualitative concept of ‘word in language L’ does not lead to interesting quantitative generalizations about texts in L, this can no more be taken to imply that the qualitative concept is to be replaced by another, ‘better’ one, than any amount of neuro- or psycholinguistic research can lead to ‘amending’ or ‘revising’ a qualitative concept. Even if we had a theoretically deducible stochastic regularity that works well with a certain concept (set of criteria) C_1 of ‘word’ but does not work at all with another set of criteria C_2 , this would not tell us that we should henceforth use C_1 instead of C_2 in our *qualitative* descriptions since the viability of a qualitative concept can only be judged relative to the qualitative theory it forms a part of. It is the qualitative delimitation of the concept that gives a stochastic statement its meaning in the first place. Observable statistical regularities about artificial constructs that have no independently discernible place

in a linguistic description (as might be the case with C_1) are virtually meaningless. As B. Mandelbrot has pointed out, qualitative and stochastic treatments of language are mutually incompatible and complementary; I would like to add that the stochastic treatment is conceptually dependent on the qualitative one but not *vice versa*.

It should be clear from the outset that none of the quantitative ‘laws’ proposed by QL so far can be expected to hold without exceptions, in striking contradistinction to the laws of fundamental physics. Given any QL ‘law’, it is always possible to artificially construct a counterexample, say, a text violating the stipulated stochastic regularity. As a matter of fact, exceptions to the inductive generalizations proposed in QL work are found anyway as soon as sufficiently large corpora of samples (usually, texts) are examined. The observations just mentioned could, and indeed should, be taken as an indication that QL does not have the same science-theoretical architecture as, say, fundamental physics, where ‘laws’ are assumed to hold without exception. Generally speaking, there are no good *a priori* reasons to believe that any ‘proper’ scientific theory must have laws in the very same sense that a small subset of the natural sciences is based on laws. QL proponents, however, would like to see the scientific apparatus of QL in complete conformity with that of the natural sciences.

It is often assumed that the inductive generalizations of QL are indeed ‘laws’ proper, if only a special kind of them, to wit, *ceteris paribus* laws that hold only when certain necessary preconditions are satisfied. However, since those necessary preconditions can – in virtue of the *ceteris paribus* restriction – not be stated explicitly and are, therefore, not specified by the law itself, the *ceteris paribus* clauses amount to no more than a trivial immunization strategy.

Thus, when one assumes, with Altmann (2002, 22), that “language laws hold only for homogeneous data” then it is difficult to avoid downright circularity, as homogeneity of data can most likely be defined only in terms of the law in question: data are homogeneous just in case the law is applicable to them.

Moreover, in the case of supposed QL laws, violations can be produced systematically and intentionally, as I already pointed out. To give but one example, given any specific candidate for a word length distribution regularity in natural language texts, it is possible to systematically construct texts that cannot be subsumed under the proposed regularity. Regularities that can be violated in an operational manner are neither laws nor *ceteris paribus*-laws (see Mott 1992, 462 for further elaboration of this point).

The reader might protest here, pointing out that there does indeed exist a well-established notion of *ceteris paribus*-law in science theory. In a similar spirit, Lehfeldt, Altmann (2002, 331; 341) hint at the oft-repeated claim that laws in *all* sciences always come with some *ceteris paribus* clause, obviously to turn down the suspicion that QL might be a second class science, when compared to fundamental physics. However, the historical source of such opinions is a misunderstanding of an important insight by Hempel formulated, e.g., in Hempel (1988). Earman, Roberts (1999) provide an excellent discussion of this point. The following quotations may serve as a summary of their argument:

Hempel’s claim is that typically a theory T of the advanced sciences will not have *any* logically contingent consequence S whose non-logical vocabulary belongs entirely to V_A [the set of ‘antecedently understood terms’ of T , P.M.]. What we can hope to derive from T are consequences of the form $P \rightarrow S$, where again S is a logically contingent sentence whose non-logical vocabulary belongs entirely to V_A and P is a “proviso” that requires the use of V_c [the set of theoretical terms first introduced with T , P.M.].¹¹

Hempel’s provisos are ... simply conditions of application of a theory which is intended to state lawlike generalizations that hold *without* qualification. Indeed, Hempel makes it explicit that his provisos are clauses that must be attached to *applications of a theory* rather than to law-statements...¹²

¹¹ Earman, Roberts (1999, 442).

¹² Earman, Roberts (1999, 444).

There is, in fact, an ongoing debate in the philosophy of science about whether a substantial notion of empirically non-void *ceteris paribus* laws can be found at all. J. Earman and J. Roberts provide a careful and painstaking survey of recent proposals on saving *ceteris paribus* laws from vacuity and argue that “not only is there no persuasive analysis of the truth conditions for *ceteris paribus* laws, there is not even an acceptable account of how they are to be saved from triviality or how they are to be melded with standard scientific methodology” (1999, 439). It is worth quoting their conclusion at some length:

There *is* a clear sense to be given to the notion of a “near-law”, i.e. a generalization that is not a strict law, but that deserves to be called a “near-law” because it is, in a precise sense, true or approximately true in almost all intended applications, because it plays the role of laws in giving explanations, supporting counterfactuals etc., and because it is clear that it makes definite claims about the world and can be confirmed or disconfirmed empirically. But, we claim, the most clear paradigms of such laws (viz. the laws of phenomenological thermodynamics) are not thought of as *ceteris paribus* laws, and statements that are thought of as *ceteris paribus* laws do not answer to this clear sense of a “near-law”. [...]

In the light of this, we wish to make the following suggestion. “*Ceteris paribus* laws” are not what many philosophers have taken them to be, that is, they are not elements of typical scientific theories that play the same kinds of roles in the practice of science that less problematic statements such as strict laws or near-laws (in the sense just defined) play. Rather, a “*ceteris paribus* law” is an element of a “work in progress”, an embryonic theory on its way to being developed to the point where it makes definite claims about the world. [...] To revive a now-unfashionable notion, “*ceteris paribus* laws” belong to the context of discovery rather than to the context of justification. [...]

If laws are needed for some purpose, then we maintain that only laws will do, and if “*ceteris paribus* laws” are the only things on offer, then what is needed is better science, and no amount of logical analysis on the part of philosophers will render the “*ceteris paribus* laws” capable of doing the job of laws (1999, 465-466).

Earman and Roberts concede that the remarks just quoted look “at first glance to be a negative judgment about the special sciences as compared with fundamental physics”. However, their intent is to reject a “misguided egalitarianism about the sciences”:

It is not “*ceteris paribus* all the way down” – *ceteris paribus* stops at the level of fundamental physics. But we are *not* physics chauvinists [...], for we deny that the mark of a good science is its similarity to fundamental physics. The concept of a law of nature seems to us to be an important one for understanding what physics is up to, but it is a misguided egalitarianism that insists that what goes for physics goes for all the sciences. The special sciences need not be in the business of stating laws of nature at all, and this blocks the inference from the legitimacy of these sciences to the legitimacy of *ceteris paribus* laws. For us, it is ironic that an effort to justify the special sciences takes the form of trying to force them into a straitjacket modeled on physics. We think this effort should be resisted, since it damages both our understanding of the special sciences and our understanding of the concept of a law of nature (1999, 472).

5. Mechanisms or metaphors?

In view of these difficulties it is natural to look for a *scientific* instead of a *science-theoretical* treatment of systems that show a certain behavior in a ‘more often than not’ fashion which is not open to a deterministic or mechanistic micro-level description. And indeed the past few decades have seen the rise of a whole bunch of scientific disciplines – theories of complexity, catastrophe, chaos, dissipative and self-organizing systems – that deal with phenomena of this kind. So it is hardly surprising to find many QL researchers using concepts like ‘attractor’, ‘self-organized criticality’ and ‘synergetic order parameter’ as background metaphors for quantitative descriptions of linguistic phenomena.

Of course, it is the received view of modern QL that these concepts are *not* used merely metaphorically: language is assumed to simply *be* a self-organizing system that functions in a way analogous to, say, self-regulating dissipative systems in chemistry. However, as Kanitscheider (1998, 23) emphasizes, the mere *claim* of analogy is not enough; it is a hypothesis that has to be *proved* on empirical grounds. Hence, transferring a formal model such as Haken's synergetics to a new domain of phenomena is tantamount to setting up a new *theory* that must be validated *independently* of previous applications of the model in other domains.¹³ In other words, the right motto should be: first set up your linguistic theory, then try to find a common denominator with theories from other fields.

A typical example of an ill-defined formal metaphor is the rather overused notion of *attractor*. Take the following statement: "Drawing on chaos theory, one supposes that theoretical linguistic units – often referred to as "emes" in theoretical language – are attractors, and that variations and changes in the unit represent shifts toward another attractor [...]. Only against this background is it possible to speak of self-organization in language, anyway" (Altmann 1996). It may well be doubted whether the attractor concept, when applied to notions of traditional linguistics, provides any additional theoretical insights or whether all we get is a vague feeling of plausibility. If a statement like "phonemes are attractors" is to have any empirical content, we must have a sufficiently elaborated mathematical concept of 'attractorhood' that may be applied to linguistic data in order to generate empirically testable hypotheses. Nothing of this kind seems to be in sight today.

It might be instructive at this point to discuss a particular example in some depth in order to shed some light on the purported *explanatory power* of the attractor metaphor. Lehfeldt, Altmann (2002) try to account for a certain sound change in Old Russian, the so-called fall of the two yer vowels.¹⁴ Their theoretical starting point is 'Menzerath's law' that, in its modern quantitative version (which is now usually called the Menzerath-Altmann Law; principal references are Altmann 1980 and Altmann, Schwibbe 1989), relates the length of a linguistic construction to the length of its constituents through an inverse proportionality, or, more generally speaking, through a power law. Applied to word length, the 'law' in its *basic form* may be written as the equation $Y = Kx^{-b}$, where x symbolizes word length as expressed in number of syllables and Y is the average number of phonemes per syllable in words of x syllables length in a given, homogeneous text. K and b are constants that may vary for different

¹³ Cf. Lees' critical comments (1959, 285ff.) on an older attempt to find a useful cross-discipline analogy between linguistics and thermodynamics based on the notion of entropy: „It is difficult to see, however, how this formal similarity between measuring the elementary message capacity of a source by partition of symbol probabilities and measuring the unavailability of heat energy by partition of atomic states can be pushed any deeper.“ (293); „To summarize, then: for good reasons, the communications engineer has been led to characterize the utility of a message source or transmission line in terms of the variety of distinct messages which it permits one to identify (with no consideration of the meaning or understandableness of the messages), and the most convenient expression for this measure involves the logarithm of a probability. For independent reasons, the physicist has been led to characterize the irreversibility of natural energy transformations in terms of a thermodynamic property of systems, the entropy, and he has shown that this property is calculable from the distribution of the particles of the system among available energy states, the expression for entropy then involving the logarithm of a probability. Therefore, the expression for selective information-content and for physical entropy are formally similar; in fact, the very same type of expression, involving the logarithm of a probability, may be used in any number of unrelated problems as a measure of degree of equidistribution.“ (295). A similar critique of useless formal analogies may be launched against modern complexity theory; as Horgan writes in a well-known popular science article (1995): "Too many simulators also suffer from what Cowan calls the reminiscence syndrome. "They say, 'Look, isn't this reminiscent of a biological or physical phenomenon!' They jump in right away as if it's a decent model for the phenomenon, and usually of course it's just got some accidental features that make it look like something."

¹⁴ The sound change took place not only in Old Russian, but in all Slavic languages, with different results in detail. Basically, two vowels going back to PIE **i* and **u* were eliminated in certain positions and merged with other vowels (in Old Russian, *e* and *o*) in all other positions.

texts. As is obvious, Menzerath's Law dictates a monotonic functional relationship between construct and constituent size. Now, in Old Russian before 1000 AD certain syllable-internal phonotactic restrictions precluded this monotonicity. The authors conclude that Menzerath's Law – in its basic form – was not operative at that time; indeed, curve-fitting leads to negative results for texts that were written before the fall of the yers. The general line of their argument makes use of a *ceteris paribus* reading of the law as criticized above: "It is important to remember right from the outset that Menzerath's law, like any other law, *holds only when the necessary preconditions for it hold*."¹⁵ Talk of 'necessary preconditions' is somehow misplaced here since there is no way to actually *specify* those preconditions except in a circular way: If the alleged law fails in a particular case of application, then it is simply *assumed* that *some* unknown precondition does not hold. The phonotactic facts of Old Russian before the fall of the yers do *not* provide any such unfulfilled precondition. They just give some independent *prima-facie*-indication that makes clear from the outset why we may not even *expect* Menzerath's Law to be applicable here. In saying "Menzerath's Law does not hold here *because* the phonotactics render a monotonic relationship impossible" the *because* must not be understood causally, but *epistemically*, as in "Mr. Lees is not ill *because* I saw him playing with his children on the street just an hour ago."

The assumption that Menzerath's Law was not operative in Old Russian before the fall of the yers has two different possible readings that should be distinguished sharply but get somehow blurred in the authors' presentation, as becomes obvious in their surprising claim (2002, 341) that during the development of Old Russian phonology Menzerath's Law never lost its force, functioning as an attractor that tries to gain control but sometimes, as the reader has to surmise, nevertheless loses its power. The two readings I announced are as follows.

- a) *Menzerath's Law, as applied to the ratio of syllable length to word length, holds without exceptions; due to interfering factors, however, it may happen that it does not directly show up in the data.* – This reading has its analogue in physics. Thus, Newton's law of gravitation knows from no exception whatsoever within the framework of classical mechanics, but if we investigate the trajectory of a feather falling in the spring air, the law does not, so to speak, 'shine through' the data because of additional complications such as movement and friction of air. Newton's theory of gravitation is not falsified by such an example, as we can make up, if only in principle, an empirical theory that accounts for the additional factors. Vector addition of the effects of these factors to the effects of Newton's

$$F = \gamma \frac{m_1 \cdot m_2}{r^2}$$
 should then yield what is actually observed. Now, the understanding just

sketched is implicit in Lehfeldt's and Altmann's talk (2002, 333) of the "possibility of anomalies and boundary conditions that disturb the monotonic direction of the curve." Indeed, a standard strategy in QL to cope with failures of alleged *ceteris paribus* laws is to assume that not all causally relevant factors have been found and accounted for in the deduction of the stochastic regularity postulated. This leads to assuming some ill-defined disturbance factor that is reflected in the derivation by means of one or more additional parameters. As the authors show, modified versions of Menzerath's Law (obtained by adding one or even two parameters to the differential equation that the Law obeys) can indeed be fitted successfully to the Old Russian data in question. However, in contradistinction to our example from physics, no independent theory of the assumed disturbance factors is available, that is, the additional parameters do not receive any empirical interpretation. All that we get, then, is a statement to the effect that adding new parameters

¹⁵ Lehfeldt, Altmann 2002, 331. All following translations from the Russian text of this article into English are mine; in the above quotation, emphasis is mine.

to a formula will improve the results of curve-fitting – a mathematical truism with no immediate linguistic implications.

- b) *Menzerath's Law, as applied to the ratio of syllable length to word length, is not operative in certain 'extremal' linguistic situations*, such as the one of Old Russian before 1000 AD. In this reading, talk of attractors seems to be more appropriate since the dynamics of a system might, under certain circumstances, be far removed from the system's attractor(s). In standard definitions of the attractor concept, however, once a system has *attained* its attractor state it will simply remain there forever. The only way, then, to explain why Old Russian before 1000 AD had left the 'Menzerath attractor' is to assume that the Old Russian language system was, during a certain period of time, determined by the effects of *another* attractor. Since attractors are abstract characterizations of the dynamics of a system, this amounts to claiming that Old Russian, at a certain stage of development, *changed* the overall look of its dynamics at least twice, losing the 'Menzerath attractor' before 1000 AD (and being forced to 'wander' to another one) and reenacting it afterwards. Of course, this sort of explanation simply shifts the burden of explanation since what we would need now is (i) a theory of the way the dynamics changes, constructing and deleting attractors in the course of time, and (ii) an explication of the presupposed 'normalcy' or 'default character' of the 'Menzerath attractor' that is implied by the 'law' terminology. Put simply, the first requirement says that *if* we claim to have *explained* the fall of the yers by pointing to the default presence of a certain sort of attractor, *then* we must also be able to explain why the default attractor vanished for a certain period of time in the first place. Requirement (ii) is even more delicate since it points to a stipulated asymmetry in the change of the overall dynamics: Why is loss of the 'Menzerath attractor' the marked alternative vis-à-vis its (re-)establishment that Lehfeldt and Altmann, if implicitly, dub a return to normality?

To sum up: None of the mechanisms one might propose in order to explain temporary 'absence' of the operation of Menzerath's Law can be backed up by anything like an empirical theory – the stipulated attractor remains but a *façon de parler*.

For Old Russian texts after the fall of the yer vowels, Menzerath's Law in its basic form can satisfactorily be fitted to the data. The authors conclude (Lehfeldt, Altmann 2002, 338): "In other words, the fall of the reduced vowels was directed at the elimination of these obstacles [for the law, PM]." Here we see an example of a *post hoc, ergo propter hoc* fallacy, that is, of an illicit causal-final reinterpretation of a merely temporal sequence of events: *After* the yers fell, Menzerath's curve could be fitted again, *therefore*, or so the argument runs, the fall of the yers *caused* or *was directed at* reenacting Menzerath's Law. We have no sound reason to take such a conclusion for granted since it is based on an unwarranted reification of the stipulated reason for the *ceteris paribus* regularity: The claim that 'Menzerath's Law holds *ceteris paribus*' effectively gets rephrased as follows: 'Menzerath's Law is a kind of "telos", a "driving force" that is somehow determined to change the dynamic structure of a language system in the course of time.'

Our result, then, is somewhat negative. No cogent explanation of the language change in question has been offered. Note that even if we had an acceptable scientific theory according to which Old Russian was forced to return to a Menzerath-compatible state, we would still stand in need of an explanation of the yer fall as such because the "goal" of reenabling Menzerath's Law could have been achieved by a host of theoretically possible ways. Indeed, Lehfeldt and Altmann reproach other purported 'explanations' of the yer fall for not providing an answer to their crucial question (2002, 328): "But why did these vowels [that is, these and not others, PM] change at all?" Of course, an answer to this question is part and parcel of the explicitly avowed main objective of the paper under discussion, viz. "to find an explanation

for the fall of the reduced vowels” (Lehfeldt, Altmann 2002, 330). The authors claim to actually have found such an explanation (2002, 342), although their formal descriptive apparatus (stating that Menzerath’s Law can be fitted to Old Russian data only after the yer fall) can not even in principle give an insight into the actual ‘mechanism’ that effected the reestablishment of Menzerath’s Law – an insight that would, amongst other things, require a complex *phonological* treatment of the Old Russian language.

The ontologizing strategy of positing attractors to explain *ceteris paribus* phenomena seems to me to be a special example of a common argumentative fallacy in QL work: A Poisson/optimization/... process can be used to *model* the phenomenon X, *ergo* there must be such a process, and the process is the searched-for explanans or mechanism; more generally: X shows a certain stochastic regularity R, *ergo* there must be an *underlying causal story* for X that *directly implies* R. Following Lees, we may say that this kind of reasoning instances an “as-if fallacy”. Lees, quoting an earlier paper by Quastler, acknowledges that certain statistical properties of music and language are equal to those which could be obtained by stochastic processes – „but it is not claimed that words grow by chance accretion of syllables, or that Mozart’s musical line is the result of random collisions“ (Lees 1959, 288).¹⁶

A famous example already discussed in detail by B. Mandelbrot, R. Lees¹⁷, G. Herdan and many others is ‘Zipf’s Law’. In recent publications Wentian Li has argued anew that “Zipf’s law is not a deep law in natural language as one might first have thought” and that “Zipf’s law does not share the common ground with other scaling behaviors”, emerging instead from ultra-general stochastic premises that hold as well for randomly generated texts (Li 1992).¹⁸ In Mandelbrot’s and Li’s interpretation, Zipf’s Law simply says that natural language texts typically behave, from a stochastic point of view, *as if* they were the output of a random character source. Naturally, this does not mean that such texts *are* such an output. Once again, the search for a mechanism ‘behind’ the stochastic regularity is determined to fail.¹⁹

6. Conclusion: Theories and models in the ‘system-determined’ sciences

Difficulties in quantitative and stochastic modeling similar to those outlined above with respect to QL arise as well in other numerically oriented branches of science. Econometrics

¹⁶ Lees characterizes the epistemic gain from quantitative linguistic models of his time by pointing out that „...the statistical behavior of words in a text, as specified by the explanatory model given, though it results from the operation of ‘known’ micro-behavior (i.e. the application of detailed grammatical rules, sociological and psychological determinants of vocabulary, etc.), could also have resulted from the operation of the probability model“ (Lees 1959, 287).

¹⁷ Lees summarizes Mandelbrot’s famous mathematical discussion on Zipf’s Law, pointing out that the law „says merely that whatever the micro-behavior may be that determines our choice of words (what we like to talk about, the grammatical constraints of our language, etc.), it results in an essentially random placement of spaces“ (Lees 1959, 287).

¹⁸ Cf. Miller’s (in)famous remarks on Zipf’s Law in his ‘introduction’ in Zipf’s *Psycho-Biology of Language* (Miller 1968): “Faced with this massive statistical regularity, you have two alternatives. Either you can assume that it reflects some universal property of human mind, or you can assume that it represents some necessary consequence of the laws of probabilities. Zipf chose the synthetic hypothesis and searched for a principle of least effort that would explain the apparent equilibrium between uniformity and diversity in our use of words. Most others who were subsequently attracted to the problems chose the analytic hypothesis and searched for a probabilistic explanation. Now, thirty years later, it seems clear that the others were right. Zipf’s curves are merely one way to express a necessary consequence of regarding a message source as a stochastic process.”

¹⁹ Lees remarks: „The fact that natural language texts are fair approximations to such random sequences shows merely that linguistic constraints, stringent though they seem to be, still permit sufficient variety in a very long text to approach the ideally random distributions. We see then that the only thing about such frequency distributions which is of immediate interest to the linguist is precisely the departures of natural language texts from the ideal distributions“ (Lees 1959, 285).

would seem to be a case in point. In a critique of probabilistic econometric modeling, R.-E. Kalman (1980, 1983) defends a methodological distinction between ‘natural’ and ‘system-determined’ sciences. The natural sciences – again with the standard example of theoretical physics – deal with laws of nature in a strict sense that hold regardless of which system is actually considered and that form the backbone of a *theory-driven* approach to empirical phenomena. System-determined sciences such as economics and engineering, on the other hand, venture a *data-driven* approach to highly system-specific regularities. For this reason, their ‘laws’ often have no validity beyond the specific sort of system to be described and may contain parameters that are neither universal constants nor liable to a general theoretical interpretation. This leads to a remarkable situation in econometrics where mathematically simple formulas with only weak theoretical motivation often turn out to be superior to sophisticated, theory based systems of differential equations when it comes to predictive capacity.

It seems to me that the situation of QL is similar. While the models developed so far do possess statistical significance, theoretical underpinnings remain vague and weak. The *methodic side* of QL research work is close in spirit and in its formal aspects to the rough-and-ready inductive generalizations of statistical modeling in the social and economic sciences, whereas its *rhetoric* is that of a super-general, if virtually non-existent, theory of complex, self-organizing systems.

The ‘big question’ that comes to mind here is whether a “third way” besides a traditional, qualitative understanding of the subject matter of linguistics and the inductive quantitative descriptions of contemporary QL (the empirical side of which rests entirely on qualitative notions in a poorly understood way) is conceivable at all. Recent contributions to the theory of complex systems suggest that qualitative-only and even functional treatments of systems may, in many scientific contexts, be both inevitable and explanatorily fruitful. J. Cohen and I. Stewart (1994) outline a theory of complex phenomena arising or ‘emerging’ from non-linear causal interaction between two or more systems whose internal dynamics differ so radically from one another that none of the attractors of the individual ‘phase spaces’ of these systems coincides with any attractor of the combined phase space arising from the interaction. In such cases, the authors argue, the resulting dynamics (which cannot be described from the point of view of the contributing systems) will develop according to simple patterns that are, in a well-defined sense, independent of the complex internal details of each of the involved systems and the specific boundary conditions of the interactions as such. The authors coin the term *complicity* for this kind of interaction. Complicity-driven dynamics can, in many cases, be subjected to a merely qualitative or functional explanation; in other words, the dynamics of the combined system is not reducible to aspects of the dynamics of the systems that form its parts.

To take a favourite example of the authors’, the extremely complex evolutionary interaction between the phase space of the internal microbiological and genetic apparatus of the higher living beings on the one hand and the behavioural phase space of the macrophysical interactions of these living beings with each other and with their natural environment has, over and over again, led to the development of *wings* the anatomy and historical morphology of which differ radically from species to species. No look at the intractable details of the evolutionary development of wings in different flying species will give us a deeper *explanation* of the overall fact that flying is reinvented by evolution again and again. Only a coarse-grained, functionally minded explanatory strategy of the “capability of flying enhances overall survival chances” will do here. No reductionism is available.²⁰ Linguistics, whether ‘qualitative’ or ‘quantitative’, possibly faces similar problems.²¹

²⁰ Cf. Lees (1959, 298): „Reduction of sentences to observational vocabulary and reduction of theories to the vocabulary of physics are usually considered to be independent; indeed, most logical empiricists have by now

Two different ways out of the science-theoretical dilemma of QL and toward a “third way” seem to be conceivable. On the one hand, some of the well-attested stochastic regularities that have been found to date might turn out to be quantitative analogues of the descriptive, non-reductive patterns of complicity as assumed by Cohen and Stewart. In this case, the role of QL laws in future linguistics would be a more mundane, modest one than hitherto assumed; qualitative and quantitative research would simply coexist and be directed at different goals and purposes. QL would not be able to find the deep, hidden mechanisms by which the evolution of linguistic communicative processes proceeds. My exemplary remarks on Zipf’s Law point in this direction.

On the other hand, if a mathematical treatment of the way qualitative linguistic entities such as words, syllables and constituent structures emerge evolutionarily is more than a self-contradictory hope, then it is precisely this mathematics of the ‘complicitary’ *qualitative* concepts we linguists live by that would have to lay the foundations for a mature Quantitative Linguistics.

The preceding remarks are not meant to be an all-or-none deconstruction of the remarkable achievements of QL. Rather, the main thrust of the criticism advanced here consists in noting that the most difficult problems of the discipline are still ahead, waiting to be solved – something most QL adherents will be willing to agree to. The positivist outlook on science that is still fashionable in QL work and the over-estimation of the paradigm of fundamental physics might be an obstacle to solving the most pressing problem, viz. that of bridging the gap between traditional and probabilistic-quantitative modes of thought in linguistics. The favorite quotation of QL, Bunge’s “every thing abides by laws” (1977, 17) is indicative of a fallacy linguists have perhaps fallen prey to just too easily. While Bunge’s dictum is devoid of sense if not taken in a down-to-earth, normative reading (“if you want to do science, try to find regularities wherever you can”), it suggests that simple laws of the sort found in certain natural sciences have to underlie each and every phenomenon of the observable world in such a way that the phenomena in question become mathematically derivable, if perhaps only “in principle”, from a small set of simple equations.²² To get rid of this admittedly enchanting idea might be a difficult, but necessary step of emancipation from an obsolete paradigm of scientific research.

References

- Altmann, G.** (1980). Prolegomena to Menzerath’s Law. *Glottometrika* 2, 1-10.
Altmann, G. (1985). Die Entstehung diatopischer Varianten. Ein stochastisches Modell. *Zeitschrift für Sprachwissenschaft* 4, 139-155.
Altmann, G. (1993). Science and Linguistics. In: Köhler, R., Rieger, B.B. (eds.), *Contributions to Quantitative Linguistics: 3-10*. Dordrecht: Kluwer Academic Publishers.

abandoned reductionism in respect to observational terms, while they still often pursue the goal of reduction of sciences to some basic science such as physics. There is, however, a great similarity between the two kinds of reduction, and there is no more reason to believe that the theoretical terms of biology will be eliminable in favor of terms of biochemistry, biophysics, or any other discipline than there is to believe that any theoretical terms in any science are eliminable at all.“

²¹ An application of Cohen’s and Stewart’s thoughts to the realm of (meta-)linguistics is hinted at in Meyer 2003. There, the complicity is proposed to arise from the interaction of systems of the following two types: (i) the neurophysiological internal organization of communicating individuals and (ii) their behavioral interaction.

²² Cf. again Horgan 1995: “Artificial life-and the entire field of complexity-seems to be based on a seductive syllogism: There are simple sets of mathematical rules that when followed by a computer give rise to extremely complicated patterns. The world also contains many extremely complicated patterns. Conclusion: Simple rules underlie many extremely complicated phenomena in the world.”

- Altmann, G.** (1996). The nature of linguistic units. *Journal of Quantitative Linguistics* 3, 1-7.
- Altmann, G.** (1999). Von der Fachsprache zum Modell. In: Wiegand, H.E. (ed.), *Sprache und Sprachen in den Wissenschaften der Gegenwart*: 294-312. Berlin: de Gruyter 1999.
- Altmann, G.** (2002). Zipfian Linguistics. *Glottometrics* 3, 19-26.
- Altmann, G., Best, K.-H.** (1996). Zur Länge der Wörter in deutschen Texten. *Glottometrika* 15, 166-180.
- Altmann, G., Erat, E., Hřebíček, L.** (1996). Word Length Distribution in Turkish Texts. *Glottometrika* 15, 195-204.
- Altmann, G., Lehfeldt, W.** (2002). Review of: Haspelmath, M. (2000). Optimality and Diachronic Adaptation. *Zeitschrift für Sprachwissenschaft* 18.2, 180-268 [includes discussion articles by several contributors]. *Göttingische Gelehrte Anzeigen* 254, 123-136.
- Altmann, G., Lehfeldt, W.** (to appear). Review of Peter Siemund (ed.): Methodology in Linguistic Typology. *Sprachtypologie und Universalienforschung* 53, 2000, Heft 1. Berlin: Akademie-Verlag, 123. S. (To appear in *Göttingische Gelehrte Anzeigen*)
- Altmann, G., Schwibbe, M.** (1989). *Das Menzerathsche Gesetz in informationsverarbeitenden Systemen*. Hildesheim: Olms.
- Balzer, W.** (1997). *Die Wissenschaft und ihre Methoden: Grundsätze der Wissenschaftstheorie. Ein Lehrbuch*. Freiburg (Breisgau): Alber.
- Balzer, W., Moulines, C.U., Sneed, J.** (1987). *An Architectonic for Science*. Dordrecht: Reidel.
- Beekes, R.S.P.** (1995). *Comparative Indo-European Linguistics. An Introduction*. Amsterdam, Philadelphia: John Benjamins.
- Best, K.-H.** (1999). Quantitative Linguistik: Entwicklung, Stand und Perspektive. *Göttinger Beiträge zur Sprachwissenschaft* 2, 7-23.
- Bunge, M.** (1977). *Treatise on Basic Philosophy. Vol. 3: Ontology I: The Furniture of the World*. Dordrecht: Reidel.
- Bunge, M.** (1995). Quality, Quantity, Pseudoquantity, and Measurement in Social Science. *Journal of Quantitative Linguistics* 2, 1-10.
- Chomsky, N.** (1978). A Theory of Core Grammar. *Glott 1*, 7-26.
- Chomsky, N.** (2000). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.
- Cohen, J., Stewart, I.** (1994). *The Collapse of Chaos. Discovering Simplicity in a Complex World*. New York: Viking.
- Earman, J., Roberts, J.** (1999). Ceteris Paribus, there is no Problem of Provisos. *Synthese* 118, 439-478.
- Grotjahn, R., Altmann, G.** (1993). Modelling the distribution of word length: Some methodological problems. In: Köhler, R., Rieger, B. (eds.), *Contributions to Quantitative Linguistics*. Dordrecht: Kluwer, 141-153.
- Haken, H.** (1978). *Synergetics*. 2nd edition. Berlin: Springer.
- Hempel, C.G.** (1988). Provisos: A Problem Concerning the Inferential Function of Scientific Laws. In: Grünbaum, A., Salmon, W. (eds.), *The Limits of Deductivism*. Berkeley, Ca.: University of California Press. 19-36.
- Horgan, J.** (1995). From Complexity to Perplexity. *Scientific American* 272, 74-79.
- Itkonen, E.** (1978). *Grammatical Theory and Metascience*. Amsterdam: Benjamins.
- Itkonen, E.** (1983). *Causality in Linguistic Theory. A Critical Investigation into the Philosophical and Methodological Foundations of 'Non-Autonomous' Linguistics*. London: Croom Helm.
- Kalman, R.-E.** (1980). A System-Theoretic Critique of Dynamic Economic Models. *International Journal of Policy Analysis and Information Systems* 4(1), 3-22.

- Kalman, R.-E.** (1983). Identifiability and Modeling in Econometrics. In: *Developments in Statistics*, vol. 4: 97-136. New York, London: Academic Press.
- Kanitscheider, B.** (1998). Vorwort: Philosophische Reflexionen über Chaos und Ordnung. In: Peitgen, H.-O., Jürgens, H., Saupe, D. *Chaos. Bausteine der Ordnung: 1-33*. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag.
- Lehfeldt, W., Altmann, G.** (2002). Padenie reducirovannykh v svete zakona P. Mencerata [The fall of the reduced vowels in the light of Menzerath's law]. *Russian Linguistics* 26(3), 327-344.
- Lees, R.B.** (1959). Review of: Apostel, L., Mandelbrot, B., Morf, A.: *Logique, langage et théorie de l'information*. Paris. *Language* 35(2), 271-303.
- Li, W.** (1992). Random Texts Exhibit Zipf's-Law-Like Word Frequency Distribution. *IEEE Transactions on Information Theory* 38(6), 1842-1845.
- Meyer, P.** (2003). *Gebrauch und Struktur. Zu den pragmatischen Grundlagen grammatischer Beschreibung*. Berlin: Logos Verlag.
- Miller, G.A.** (1968). Introduction. In: Zipf, G.K. (1968), iii-x.
- Mott, P.** (1992). Fodor and Ceteris Paribus Laws. *Mind* 101, 333-346.
- Stephan, A.** (1999). *Emergenz. Von der Unvorhersagbarkeit zur Selbstorganisation*. Dresden: Dresden University Press.
- Suppes, P.** (1998). Pragmatism in Physics. In: Weingartner, P., Schurz, G., Dorn, G. (eds.), *The Role of Pragmatism in Contemporary Philosophy*. Wien: Hölder-Pichler-Tempsky.
- Weizsäcker, C.F.v.** (1985). *Aufbau der Physik*. München: Hanser.
- Wimmer, G., Köhler, R., Grotjahn, R., Altmann G.** (1994). Towards a Theory of Word Length Distribution. *Journal of Quantitative Linguistics* 1, 98-106.
- Wimmer, G., Altmann, G.** (1994). The theory of word length: some results and generalizations. *Glottometrika* 15, 110-129.
- Wimmer, G., R., Altmann, G.** (2003). Unified Derivation of some Linguistic Laws. In: Altmann, G., Köhler, R., Piotrowski, R.G. (eds.). (to appear) *Handbook of Quantitative Linguistics*. Berlin: de Gruyter.
- Zipf, G.K.** (1968). *The Psycho-Biology of Language. An Introduction to Dynamic Philology*. Cambridge/Mass.: M.I.T. Press. 2nd printing.